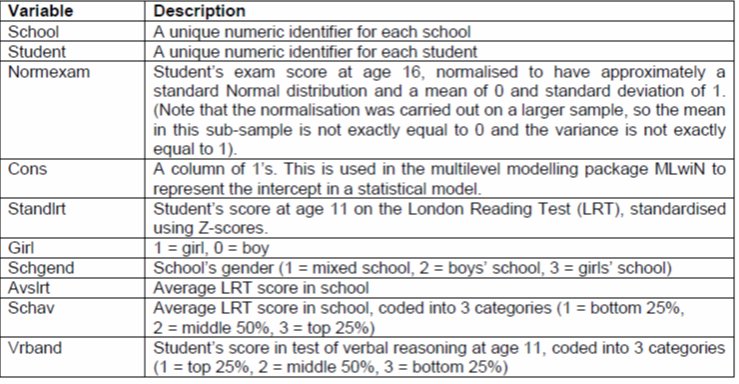
DT228A/B Probability and Statistical Inference

# Week 3 Exercise

1. Load/open your regression dataset (regression.sav)

This dataset comprises a sample of 4,059 young people (aged 16) selected from 65 difference secondary schools from six inner London Education Authorities.This is a sub-sample from a much larger study undertaken by Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., et al. (1993) A multilevel analysis of school examination results, Oxford Review of Education, 19, pp. 425-433. The dataset has been specifically prepared to accompany Rasbash, J. et al. (2005) A User’s Guide to MLwiN 2.0 (Bristol, Centre for Multilevel Modelling).



1. To read in the file we need to do the following:

library(foreign)

#Read in the file

regression <- read.spss("regression.sav", use.value.labels=TRUE, max.value.labels=Inf, to.data.frame=TRUE)

#Setting the column names to be that used in the dataset

colnames(regression) <- tolower(colnames(regression))

1. We are interested in the following variables:

Normexam – students score age 16

Standlrt – students score at age 11 on the LRT

Review these for normality

1. Investigate the correlation between Normexam and Standlrt.
2. Using Survey.dat
   * Investigate the normality of the following:
     1. perceived stress (tpstress)
     2. positive affect (tposaff)
     3. negative affect (tnegaff)
     4. life satisfaction (tlifesat)
     5. self-esteem (tslfest)
3. Investigate the following correlations:
   * Perceived stress and self-esteem
   * Perceived stress and positive affect
   * Positive affect and Life Satisfaction
   * Negative affect and Self-esteem